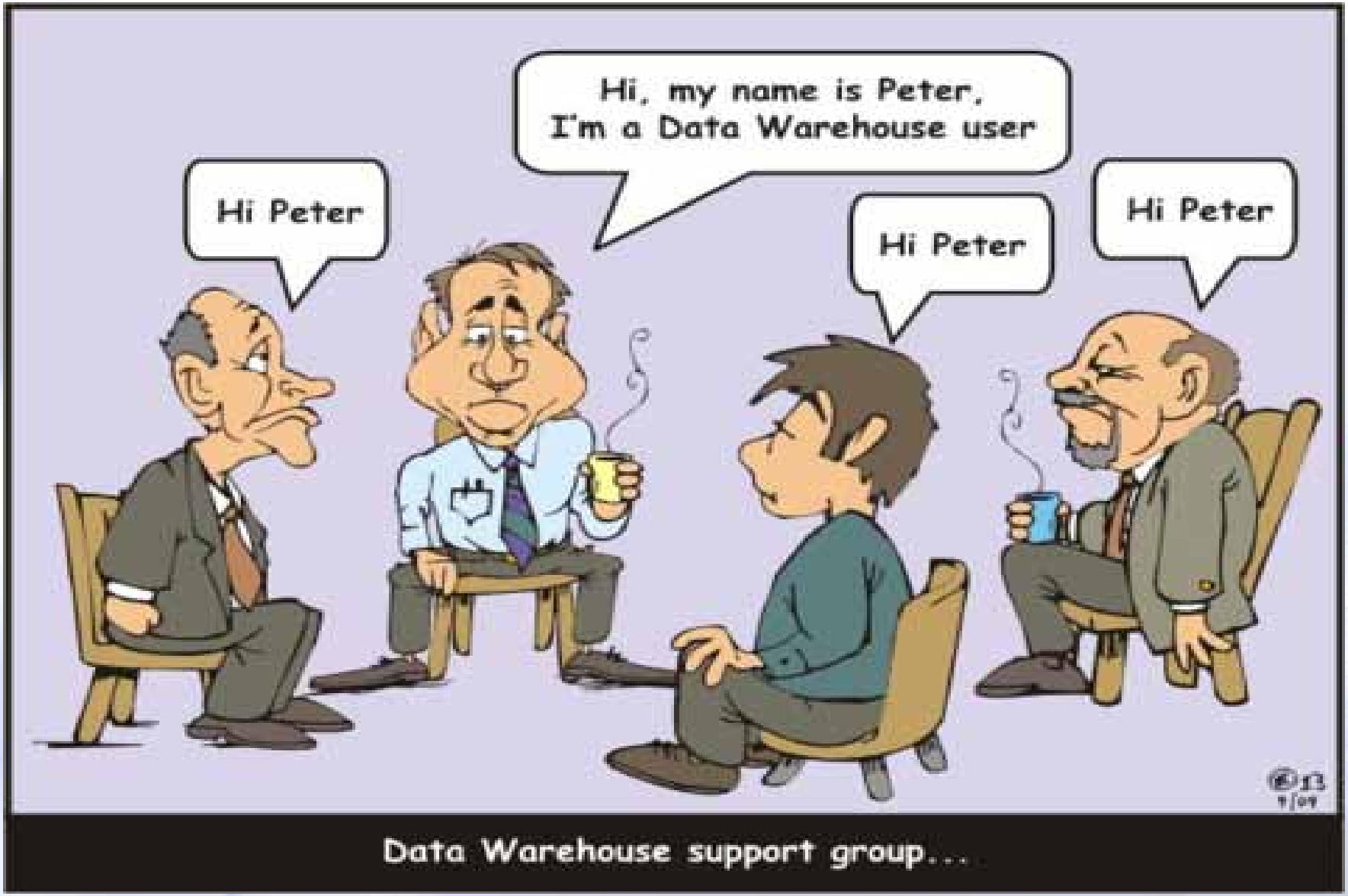
Evergreen State P-20 Program

"What is a Data Warehouse"

Executive Steering Committee Presentation

April 21, 2011

Beware - Data Warehouse Usage is Addictive



A data warehouse is a consolidated view of your enterprise data, optimized for reporting and analysis.

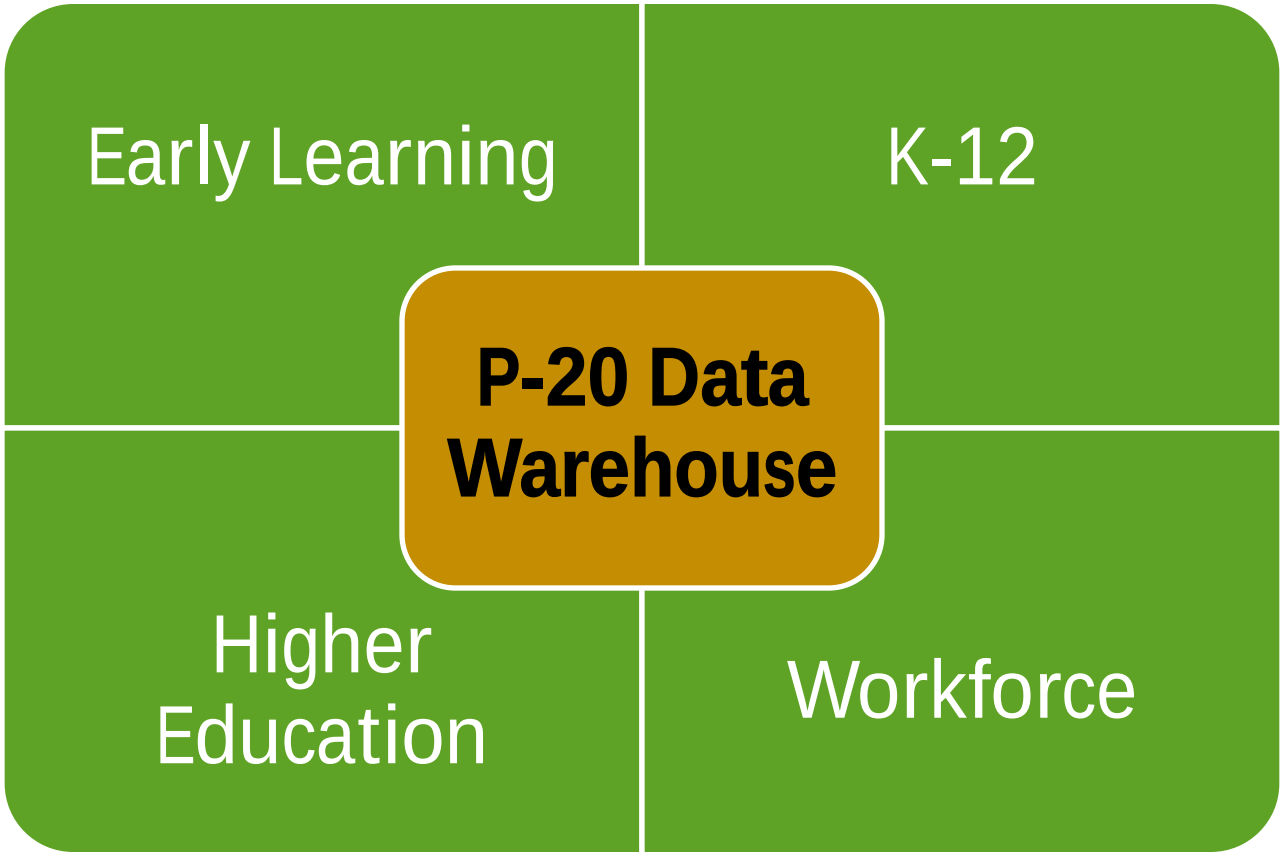
Goal of the P-20 Data Warehouse

Provide a single source of TRUTH (valid, reliable, consistent , and compliant information) for P-20 information *spanning all* sectors (Early Learning, K-12, Higher Education, & Workforce) that will empower the P-20 User Community to be:

Informed, increasing the context, relevance, and confidence of the information they need to make better decisions.

Engaged, with a more dynamic and interactive experience, making it easier to contribute and collaborate.

P-20 DW Spans All Four Sectors



P-20 Data Warehouse Primary Objectives

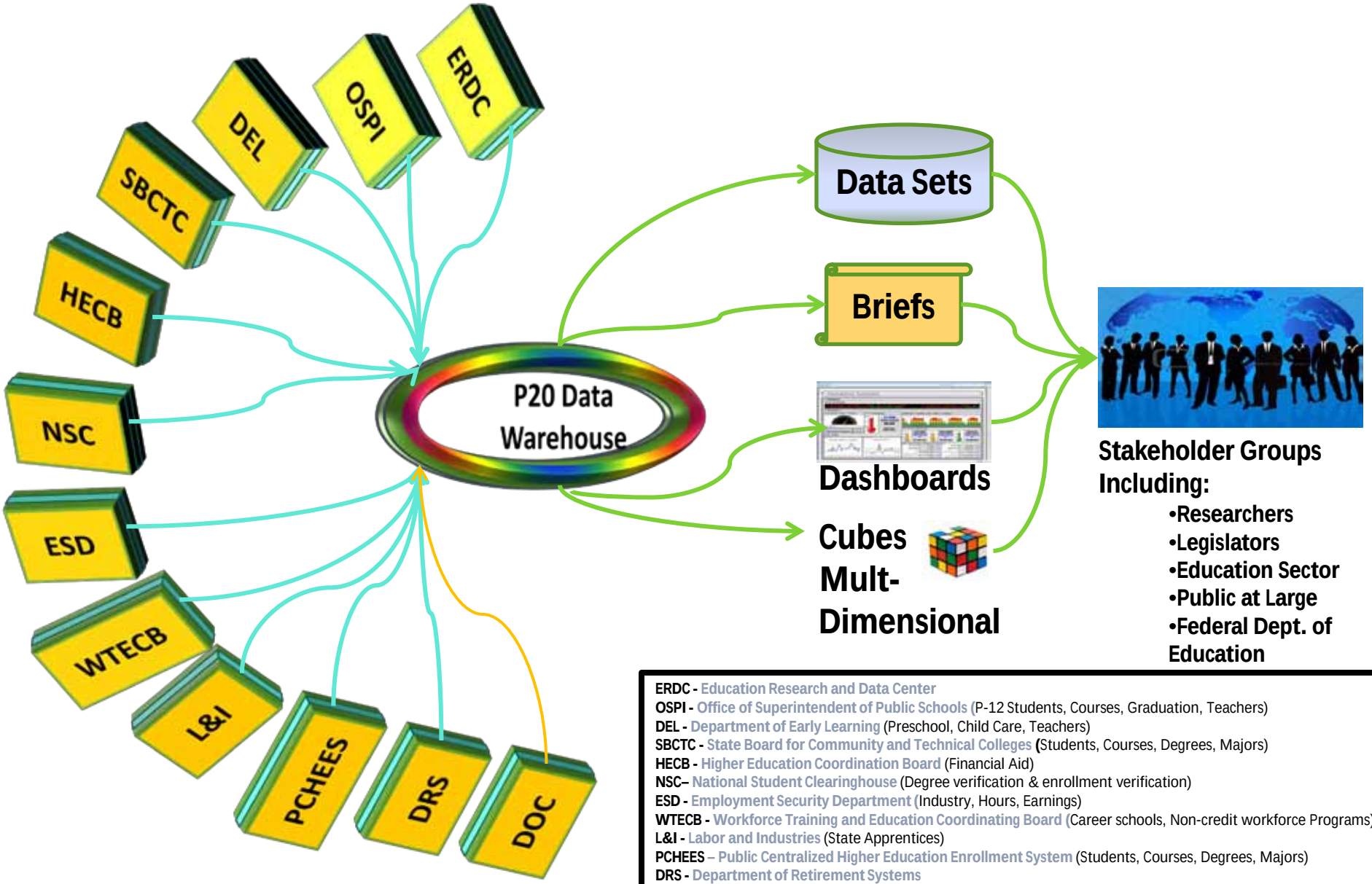
The P-20 Data Warehouse must meet the following two high level objectives*:

1. **Data Warehouse Environment**. The construction of a structured P-20 data environment will include a data inventory spanning systems, the development of a P-20 data dictionary, and the implementation of a data warehouse with a variety of data marts designed to support research and reporting. The goal is to efficiently generate research datasets and summary information and to help ensure data quality.

2. **Interoperability.** The goal is to facilitate the efficient and standardized exchange of data between the P-20 Data Warehouse and contributing data systems.

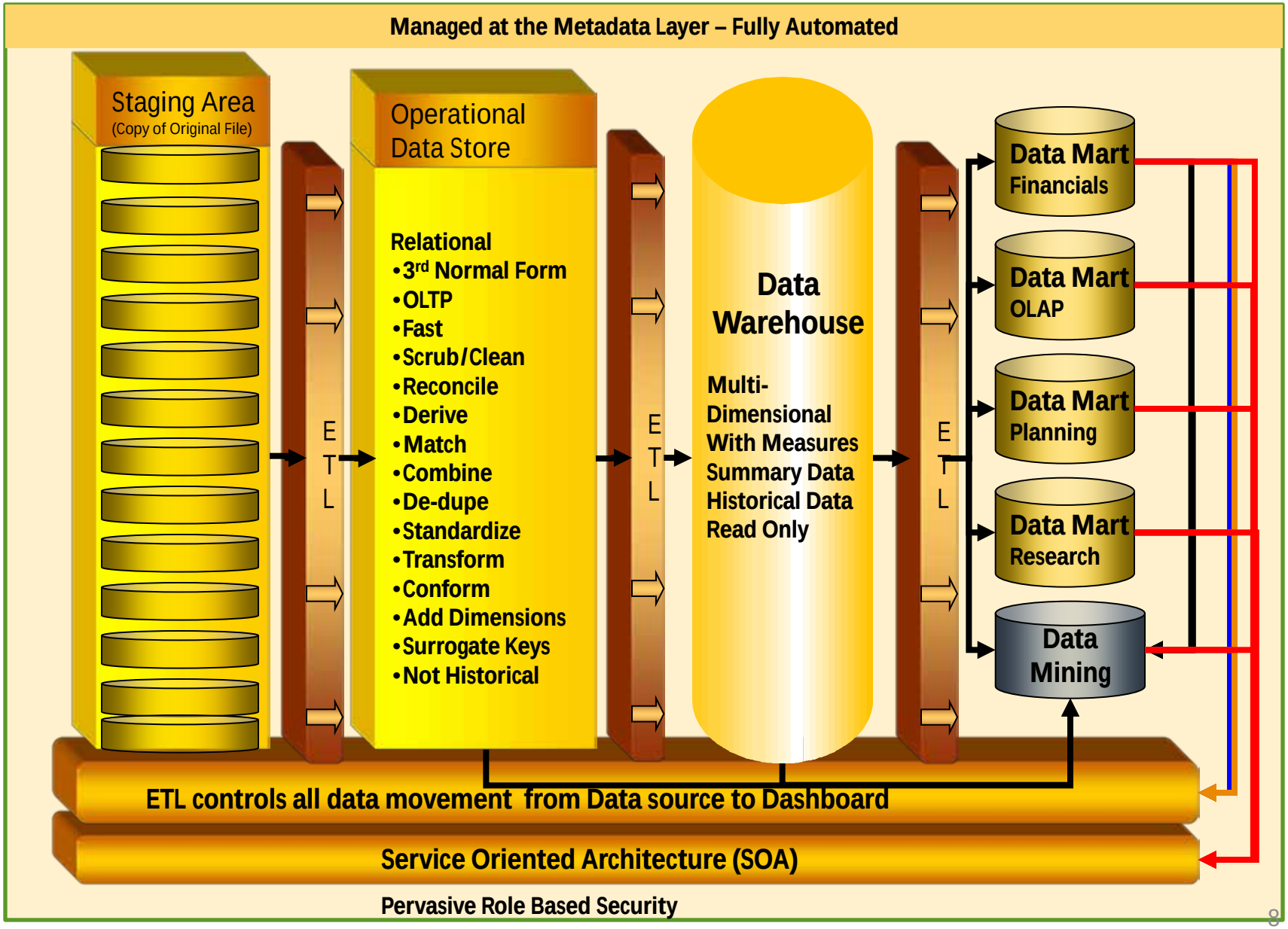
As defined in the WA State ARRA Grant Submission

P20 High Level Conceptual Diagram



ERDC - Education Research and Data Center
OSPI - Office of Superintendent of Public Schools (P-12 Students, Courses, Graduation, Teachers)
DEL - Department of Early Learning (Preschool, Child Care, Teachers)
SBCTC - State Board for Community and Technical Colleges (Students, Courses, Degrees, Majors)
HECB - Higher Education Coordination Board (Financial Aid)
NSC– National Student Clearinghouse (Degree verification & enrollment verification)
ESD - Employment Security Department (Industry, Hours, Earnings)
WTECB - Workforce Training and Education Coordinating Board (Career schools, Non-credit workforce Programs)
L&I - Labor and Industries (State Apprentices)
PCHEES – Public Centralized Higher Education Enrollment System (Students, Courses, Degrees, Majors)
DRS - Department of Retirement Systems
DOC – Department of Corrections (Inmate entries, exits, education)

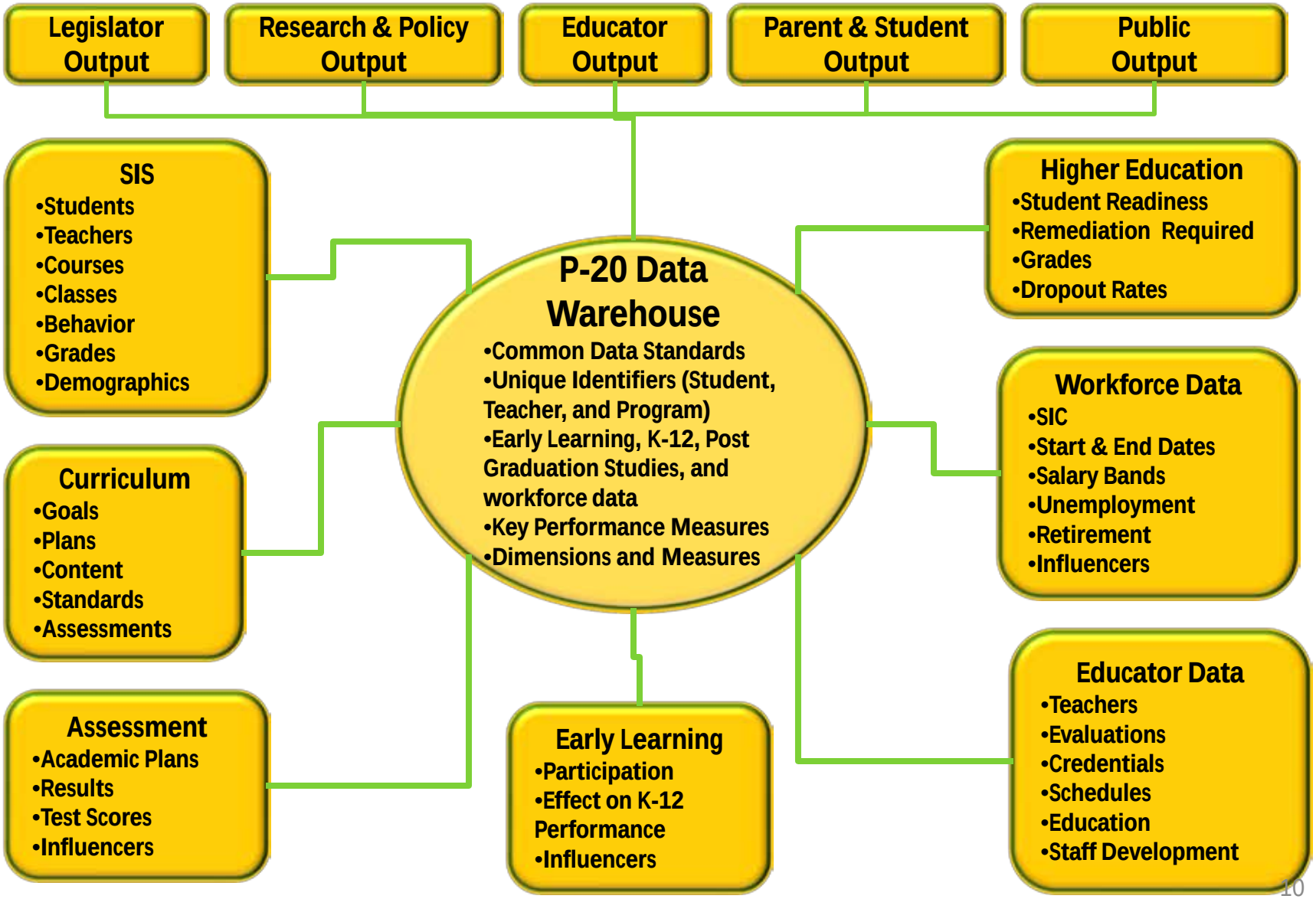
P20 Data Warehouse



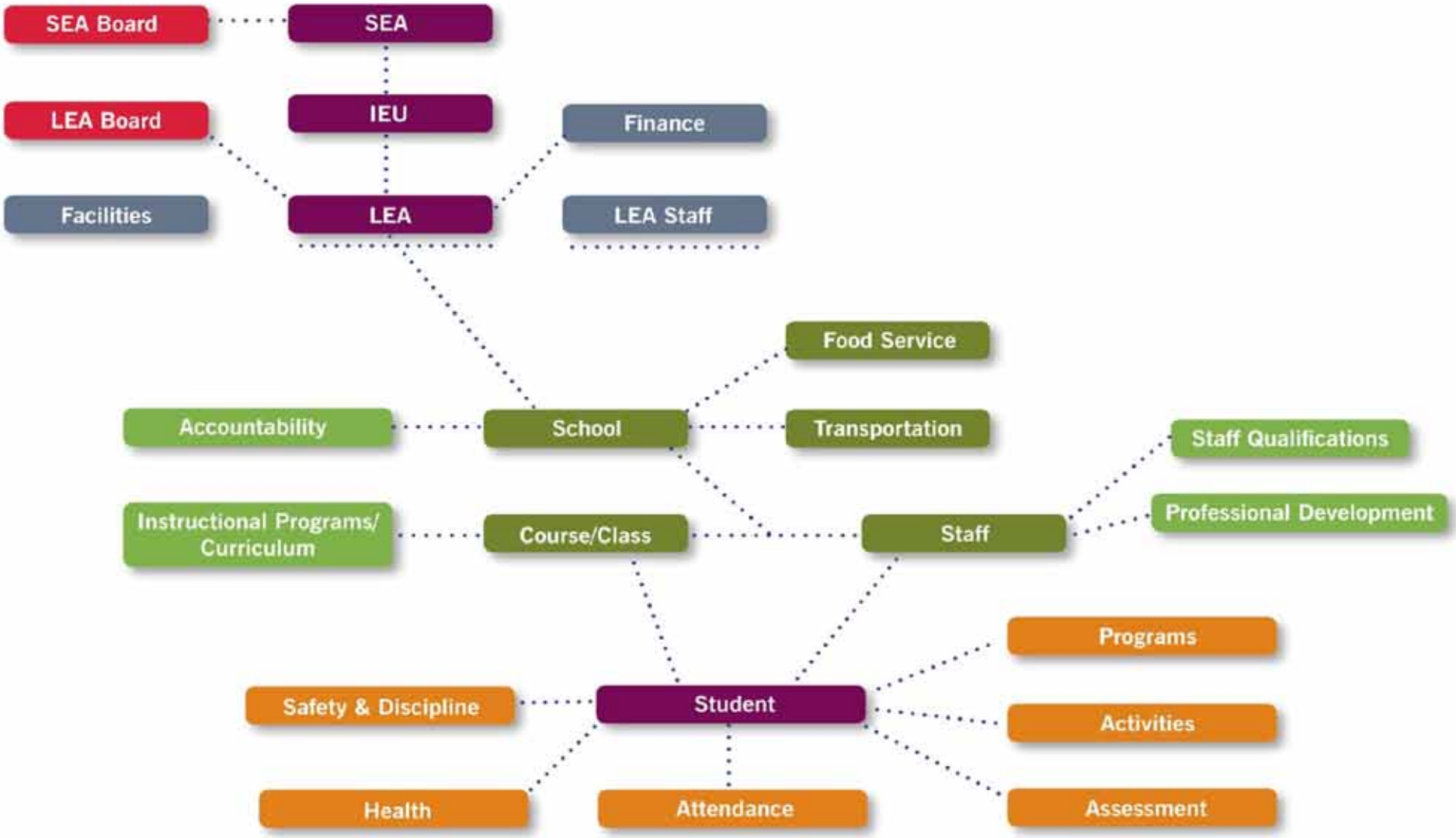
Information Hub

The information hub is a combination of two data warehousing tools managed at the metadata layer. The ETL tool is used for large data transfers and where complex transformations are required. The SOA tool is used to communicate message sized data in near real time. Metadata is shared within the ETL tool metadata manager.

P-20 DW – Example Subject Areas



NEDM: National Education Data Model



P-20 Data Warehouse Guiding Principles

1. The P-20 Data Warehouse shall be operational and useable no later than June 30th , 2013.
2. The P-20 Data Warehouse shall meet the longitudinal data needs of its Sponsors and Stakeholders.
3. The P-20 Data Warehouse shall use industry best practices to achieve its goals and outcomes.
4. The P-20 Data Warehouse is not a "Critical" State of Washington Application under disaster recovery purposes.
5. The P-20 Data Warehouse shall be a hardened environment that protects sensitive information from unauthorized access.
6. Subject to the Implementation Study findings, the P-20 Data Warehouse shall be a traditional data warehouse with Staging Area, Operational Data Store (ODS), Data Warehouse (DW), and Data Marts.
7. The Presentation Layer of the P-20 Data Warehouse may consist of a Business Intelligence capability that is accessed through a web portal using Role Based Security.
8. The P-20 Data Warehouse shall contain personal identity information, social security numbers and other information necessary to uniquely identify individuals in support of implementing a unique identifier that will allow an individual to be tracked across all sectors from early learning to the workforce.
9. The P-20 Data Warehouse shall utilize State staff as much as practical, in the design and implementation of the data warehouse, to maximize their understanding of the creation, operation and maintenance of the P-20 Data Warehouse and to minimize the risks associated with handoff of the system into the production environment.

Organizational Change

The State of Washington may need to make changes:

- ***Culturally, stakeholders may need to shift from reporting data for compliance purposes to using data*** **to guide education decisions, especially those focused on improving teaching and learning.**
- ***Politically, policymakers need to ensure that educational institutions and other relevant data contributors,*** **share student-level data — while protecting student confidentiality — to improve student achievement.**
- ***Organizationally, states need to create governance structures to ensure the effective and appropriate*** **collection and use of high-quality longitudinal data, especially as data are shared across agencies and other traditional boundaries.**
- ***Financially, states need to continue to invest in the development, maintenance and growth of their*** **education data systems, including helping stakeholders learn how to use the information produced by the systems.**

A Different View of the Data

Historically, student information was viewed within sectors.

***Longitudinal data is viewed across sectors* in the form of pipelines.**

People flow across, into, and out of pipelines over time.

Pipeline views allow you to follow cohorts as they travel through these pipelines, leave and re-enter.

- **A person may be a student at a higher education institution, leave to participate in the workforce and return again to continue higher education.**

The act that a person is performing at any given place and time is described as a role.

A person must have one to many roles over time

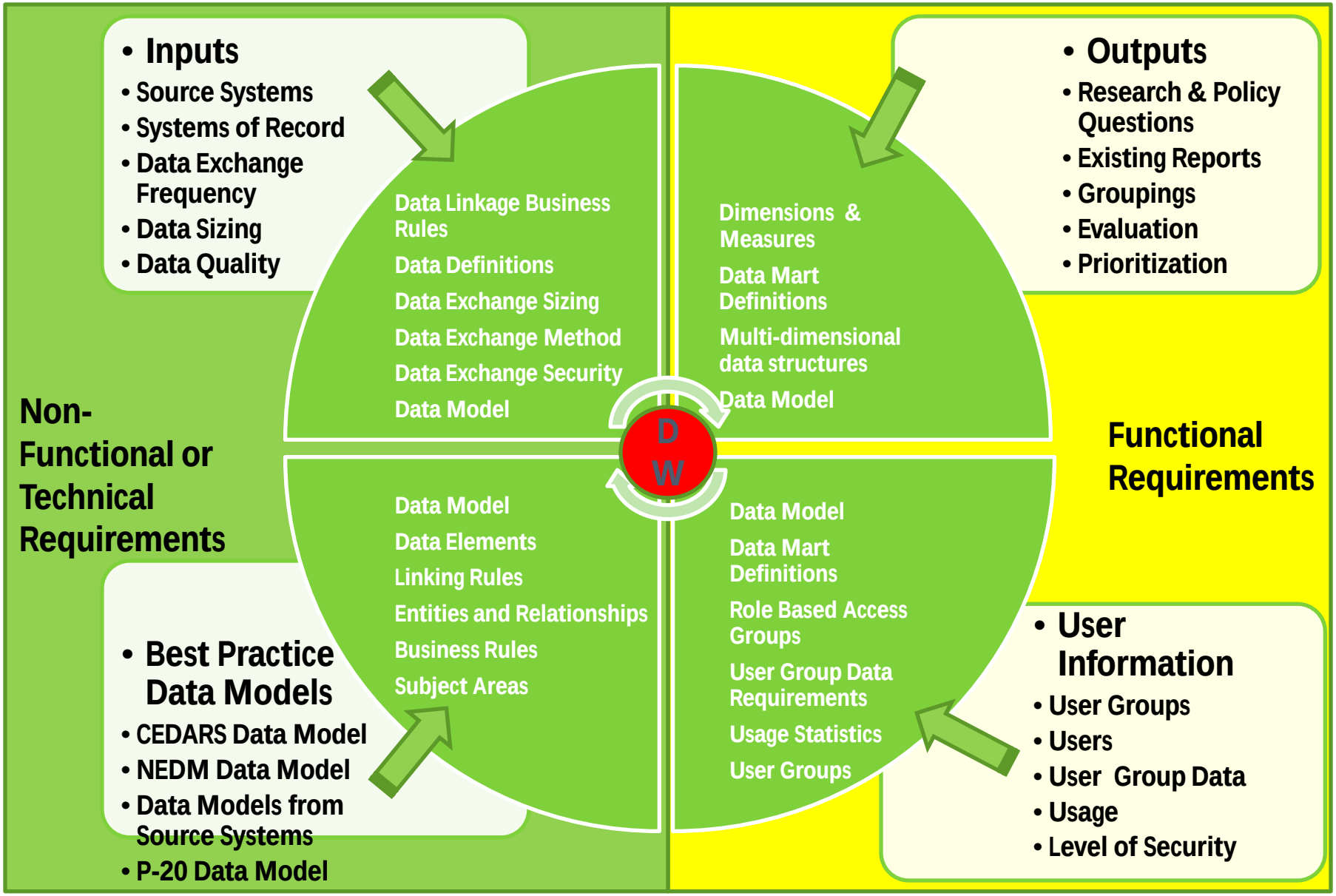
- **Early Learning Student, K-12 Student, Higher Education Student, Teacher, Workforce participant**

Defining the Data Warehouse

Before the Implementation Study team arrives, we are documenting the information needed as input to that process. Many of these input documents will continue to be defined during the Implementation Study.

- **Functional Requirements**
- **Non-Functional Requirements**
- **Input system Data Owners, Data Stewards, Data Custodians and Data Users**
- **User Information**
- **Estimated Size and Growth Rate of the DW**
- **Conceptual Architecture**
- **State of Readiness to implement and host the DW**

Defining the Data Warehouse



THE END